

# FoodAPS Retail Environment Study Data

## Documentation

**July 22, 2014**

**Project Title:** Food Acquisition and Purchase Survey Geography Component (FoodAPS-GC)

**USDA Agreement Number:** 58-5000-1-0051

**ERS Technical Contact:** Michele Ver Ploeg

**Submitted by:**

Parke Wilde, Friedman School of Nutrition Science and Policy, Tufts University

Joseph Llobrera, Friedman School of Nutrition Science and Policy, Tufts University

Note: Some of the data sources are still being updated. The new data will promptly replace older data sources mentioned in this document. This draft codebook shows users what data have been collected and will be available for research. Updates will be posted as they are available, and each new post will include an indication of the date that the update was made.

# 1. Introduction

The **Food Acquisition and Purchase Survey Geography Component (FoodAPS-GC)** investigates how the local food environment influences food spending patterns in the United States. FoodAPS-GC was developed in response to intense interest in access to affordable healthy food and the potential effects of the food retail environment on food consumption.

This project studies the same counties selected for the **National Household Food Acquisition and Purchase Survey (FoodAPS)**, a major food expenditure survey being conducted by USDA's Economic Research Service (ERS). In 2012-13, FoodAPS collected food expenditure data from approximately 4,800 households in 50 primary sampling units (PSUs), which are counties or groups of counties, and 400 secondary sampling units (SSUs), which are census block groups or blocks.

FoodAPS-GC involves two USDA/ERS cooperative agreements. An agreement with the Friedman School of Nutrition Science and Policy at Tufts University focuses on characterizing the food environment in the 50 PSUs. An agreement with the University of Illinois focuses on data describing food prices in the retail environment, and this agreement also oversees a program of extramural research using FoodAPS data to study the effect of food environments on food spending and other food behaviors.

In addition to providing the FoodAPS survey data, this project is expected to be able to provide two types of data on food retail environments:

1. **FoodAPS Retail Environment Study Data.** These data describe the presence of food retailers and other food sources in the 50 PSUs that were sampled for FoodAPS. For example, for each block group in the PSUs, these data provide the number of large food retailers (supermarkets, supercenters, and large grocery stores) and farmer's markets at selected distances from the population-weighted block-group centroid.<sup>1</sup>
2. **Food Price and Availability Data.** These data, based on point of sales (POS) data from IRI, will describe prices and availability for particular food and beverage categories in retailers within the 50 PSUs. For example, for particular retailers in a PSU, these data may provide an index showing whether the retailers offer particular categories of fruits and vegetables, and, if so, an index of the price levels.

This **Documentation** file provides data documentation for the first of these two types of data on food retail environments. A separate documentation file will cover the second type of data, the Food Price and Availability Data, which is still in development. The main distinction between the two data sources is as follows: (1) the FoodAPS Retail Environment Study Data provide comparatively simple counts, showing the presence of a wider variety of retailers and other food sources, while (2) the Food Price and Availability Data will provide more specific information about within-store prices and product availability for selected retailers or groups of retailers.

---

<sup>1</sup> These data are available at the block group, census tract, and county level. The data are intended for researchers who seek to use more aggregated food environment measures and do not need item-level or store-level data. Researchers who seek to use different distance definitions, or to classify retailers and other food sources differently from the conventions used here, may request more information from ERS through the FoodAPS email address (FoodAPS@ers.usda.gov). A set of measures of the distance between sampled households and visited food retailers will also be available in the FoodAPS survey data.

This documentation file is designed for analysts who have or expect to receive access to the FoodAPS Retail Environment Study Data directly from USDA. Typically, these analysts may be interested in the effects of food retailers and other food sources on food spending or other outcomes described in FoodAPS. Such analysis requires knowing more detailed information about the residential location of survey respondents than may be provided in a public use data file, because respondent confidentiality must be protected. Hence, USDA has developed a procedure for qualified researchers to access the data through a secure data enclave maintained by the National Opinion Research Center (NORC), under provisions that assure non-disclosure of confidential information.<sup>2</sup>

The FoodAPS Retail Environment Study Data are provided in four data files corresponding to three levels of geographic aggregation:<sup>3</sup>

- **County-level file.** This file includes information on the total and population-normalized count of food retailers and other food sources in PSU counties. It also includes county-level population counts and socioeconomic indicators. Policy variables related to school meals, sales tax rates, and Supplemental Nutrition Assistance Program (SNAP) administrative options are also provided at the county or state level. Sources include proprietary business databases, government data, and information from private organizations. The file includes 108 counties, which comprise the 50 PSUs of the FoodAPS survey.

---

<sup>2</sup>Further information is available from USDA ERS: <http://www.ers.usda.gov/data-products/foodaps-national-household-food-acquisition-and-purchase-survey/>.

<sup>3</sup> Due to differences in sampling frame, the block group-level data on restaurants exists as a separate file distinct from the rest of the block group-level data. Geographic identifiers can be used to link the files.

- **Tract-level file.** This file includes information on the presence of food retailers and other food sources in and around PSU block groups, aggregated to the census tract level. It also includes census tract-level population counts, socioeconomic indicators, and other measures of food access. Sources include proprietary business databases, government data, and information from private organizations. The file includes 11,854 tracts, which comprise all the tracts in the 108 FoodAPS counties.
- **Main block group-level file.** This file includes information on the presence of food retailers and farmers markets in and around all block groups in the PSU counties, population counts, and socioeconomic indicators. Sources include proprietary business databases, government data, and information from private organizations. The file includes 33,947 block groups, which comprise all the block groups in the 11,854 census tracts.
- **Block group-level restaurant file.** This file is based originally on InfoUSA data on the location of eating places (restaurants) in the PSU block groups where FoodAPS households reside.<sup>4</sup> Restaurants were classified as limited-service or full-service restaurants. Limited-service restaurants are establishments primarily engaged in providing food services where patrons generally order, select, and pay for items before eating. Food and drink may be consumed on premises, taken out, or delivered to the customers' location. Full-service restaurants are establishments primarily engaged in providing food services to patrons who order and are served while seated, and then pay after eating. The file includes 748 block groups.

---

<sup>4</sup> Based on Census 2010 geographic boundaries.

The four files include identifiers that allow analysts to easily combine them. This includes standardized FIPS identifiers for census block group, tract, county and State, as well as identifiers for FoodAPS counties and their PSUs. Census 2000 geographic boundaries were used during the sampling stage of FoodAPS, while Census 2010 boundaries were used at the time the analysis files for this project were created. The FoodAPS household geocode file (**hhgeocode**) contains geographic identifiers for both the 2000 and 2010 census boundaries that allow analysts to match data based on either set of boundaries.

The block group-level files offer the greatest degree of geographical disaggregation. In general, the Census Bureau defines block groups such that the population is approximately 1,500 people, with population usually ranging between 600 to 3,000 people.

The remainder of this Documentation file is organized as follows:

- **Section 2** describes the data sources, including government data sources, proprietary business databases, and information from private organizations.
- **Section 3** describes the main block group-level file and explains the system of prefixes and suffixes used to name variables.
- **Section 4** describes the block group-level restaurant file and explains the system of prefixes and suffixes used to name variables.
- **Section 5** describes the more aggregated tract-level file and county-level file and explains the system of prefixes and suffixes used to name variables.

The **Variable List-GC** contains coding for the specific variables in each file.

## 2. Data Sources

**Census 2010:** Census 2010 is the most recent national census of the United States, providing detailed counts and characteristics of the US population at all three levels of geography used in the FoodAPS data.

<https://www.census.gov/2010census/data/>

**American Community Survey:** The American Community Survey (ACS) is an ongoing statistical survey by the U.S. Census Bureau that collects demographic, housing, social, and economic information previously contained in the long form of the decennial census. Variables and analysis were based on the public use microdata sample (PUMS). Data are based on the 5-year average for 2008-2012.

[http://www.census.gov/acs/www/data\\_documentation/public\\_use\\_microdata\\_sample/](http://www.census.gov/acs/www/data_documentation/public_use_microdata_sample/)

**TDLinx/STARS:** FoodAPS-GC includes retailer data based originally on Nielsen TDLinx and FNS Store Tracking and Redemption System (STARS) sources. Nielsen TDLinx is a store/outlet-level database of retailers selling consumer packaged goods, including food. USDA's Food and Nutrition Service maintains the STARS to monitor and manage SNAP-authorized retailers. ERS merged data from the two databases to provide information on the location and type of food retailers in 2012. These data were then used to provide counts of food retailers at selected distances for all block groups in the PSU counties. Food retailers are classified as large retailers (supermarkets, supercenters, and large grocery stores) or smaller retailers (convenience stores, pharmacies, gas stations, dollar stores, small and medium grocery stores, and specialty food retailers).

**USDA Farmers Market Directory:** The Agricultural Marketing Service (AMS) of USDA maintains a listing of over 8,000 farmers markets throughout the United States. Farmers markets

are defined as "two or more farm vendors selling at a common direct retail outlet at the same physical location on a recurring basis". The listing is based on self-reported information provided by vendors in 2012, including the locations, operating times, product offerings, and accepted payment methods for the listed farmers markets.

<http://www.ams.usda.gov/AMSV1.0/farmersmarkets>

**InfoUSA restaurant database:** Restaurant data in the block group- and census tract-level files are based originally on InfoUSA's proprietary business listing of eating places. The database includes information on the street address and InfoUSA SIC industry codes for eating places in early 2012.

<http://www.infousa.com>

**Emergency Food System Data:** *Feeding America*, the umbrella organization for the charitable emergency food system, conducts its *Hunger in America* study every few years to gain insight into charitable food distribution in the US. Data from the most recent study are expected in summer 2014 and will provide information on the location of food banks, food pantries, and soup kitchens, volume of food distribution, and hours of operation.

**USDA/ERS Food Access Research Atlas:** The atlas provides a spatial overview of food access indicators for low-income and other census tracts in 2010 using different measures of supermarket accessibility. Measures and indicators of food retail access at the census tract level can be mapped using the online mapping tool at the census tract level or downloaded as a spreadsheet.

<http://www.ers.usda.gov/data-products/food-access-research-atlas.aspx>

**County Business Patterns:** Restaurant data in the county-level file are based on the 2012 County Business Patterns database. The County Business Patterns database is conducted by the

U.S. Census Bureau and provides information on the number, employment, and payroll size of business establishments at the national, State, county, metropolitan area, and ZIP code levels.

<http://www.census.gov/econ/cbp/>

**Common Core of Data (CCD):** The CCD is a program of the U.S. Department of Education's National Center for Education Statistics that annually collects fiscal and non-fiscal data about all public schools, public school districts and state education agencies in the United States. The county-level file includes descriptive information about the total number of students enrolled and the number of students eligible for free and reduced price lunch in School Year 2010-11.

<http://nces.ed.gov/ccd/>

**Classification of Laws Associated with School Students (CLASS):** The National Cancer Institute uses a scoring system to classify state laws as they compare to national standards and recommendations for physical education (PE) and nutrition. The criteria for the scoring system are based on current public health research and national recommendations and standards for PE and nutrition in schools. These data are available for elementary, middle, and high schools. The most recent data are based on laws in effect as of December 2012.

<http://class.cancer.gov>

**State Sales Tax on Soda and Snack Foods:** Bridging the Gap is a Robert Wood Johnson Foundation funded research program focused on policies and environmental factors influencing diet, physical activity and obesity among youth, as well as youth tobacco use. Information on sales tax rates for each of the 50 states and the District of Columbia for sodas and selected snack products sold through grocery stores and vending machines is based on codified law as of January 2012.

[http://www.bridgingthegapresearch.org/research/sodasack\\_taxes/](http://www.bridgingthegapresearch.org/research/sodasack_taxes/)

**SNAP Policy Database:** The SNAP Policy Database provides information on state-level SNAP policy options. The database includes information on eligibility criteria, recertification and reporting requirements, benefit issuance methods, availability of online applications, use of biometric technology (such as fingerprinting), and coordination with other low-income assistance programs. Data from the most recent month with complete data (May 2010) are provided.

<http://www.ers.usda.gov/data-products/snap-policy-database.aspx>

### **3. Main Block Group-Level File**

The variable names in the main block group-level file follow conventions to indicate their data source and meaning. Furthermore, the individual variable names are described in the Variable List-GC file.

Many variables describe counts of retailers or other food sources within selected distances from the population-weighted block group centroid. Block groups in urban and densely-populated areas have a small land area, so the population-weighted block group centroid may be treated as the location of the block group's residents. Distances to retailers or other food sources are computed from the population-weighted block group centroid. In some rural block groups, much of the population may reside in just one section of the block group, in which case the population-weighted block group centroid still represents household residential locations fairly well.

Furthermore, in the United States, most poor and non-poor rural residents have access to an automobile. Yet, it should be noted that some rural block groups have a large land area or a population that is not concentrated in one area, in which case, the population-weighted block group centroid may not represent residential locations very well.

These distances from the population-weighted block group centroid are indicated by a suffix to the variable name (Table 1). For each distance, a variable’s prefix indicates the type of variable (Table 2). For variables that indicate the presence of particular retailer types or other food sources, the middle part of the variable name states the name of the food source (Table 3). The meaning of the variable name comes from the prefix, the middle, and the suffix together. For example, for each block group, the variable *N\_Supermarket\_d1* records the non-negative integer number of supermarkets/supercenters/large grocery stores located less than or equal to 0.5 miles from a block group centroid.

**Table 1.** Variable name suffixes indicate distances from population-weighted centroids.

<b>Suffix</b>	<b>Maximum Distance</b>	<b>Full Description</b>
<b>_d1</b>	0.5 miles	less than or equal to 0.5 miles
<b>_d2</b>	1 mile	greater than 0.5 miles and less than or equal to 1 mile
<b>_d3</b>	10 miles	greater than 1 mile and less than or equal to 10 miles
<b>_d4</b>	20 miles	greater than 10 miles and less than or equal to 20 miles

**Table 2.** Variable name prefixes indicate the type of variable.

Prefix	Full Description
N_	A count of food sources within specified distances
LABG	A binary indicator variable for low access block groups, based on not having any supermarket/supercenters/large grocery stores within specified distances (1 = low access block group; 0 = not a low access block group).
LABGpop	A non-negative integer variable showing the number of persons living in a low access block group, using specified distances in the definition of low access.  This variable is coded as missing for block groups that are not low access.
LABGhunv	A non-negative integer variable showing the number of housing units with no vehicle access in block groups that qualify as limited access block groups, using specified distances in the definition of low access. This variable is coded as missing for block groups that are not low access.

**Table 3.** Variable name descriptions for retailer types and farmers markets.

Name	Full Description
<b>Supermarket</b>	Supermarkets, supercenters, and large grocery stores with annual sales greater than \$2 million. Based on the union of TDLinx and STARS retailer data.
<b>NonSupermarket</b>	Small food retailers. This includes stores in STARS not considered supermarkets, such as smaller grocery stores with annual sales < \$2M, convenience stores, pharmacies, dollar stores, gas stations, and specialty food retailers (bakeries, fruit and vegetable stores, meat stores, and seafood stores).
<b>FarmersMarket</b>	Farmers markets. Farmers markets are defined as "two or more farm vendors selling at a common direct retail outlet at the same physical location on a recurring basis."
<b>SnapFMarket</b>	Farmers markets accepting SNAP.

Note: The files named BGMain denote the most current version of the main block group file.

#### 4. Block Group-Level Restaurant File

The variables in the block group-level restaurant file describe counts of eating places (restaurants) within selected distances from population-weighted block group centroids. This file contains only those block groups in which FoodAPS households reside.<sup>5</sup> In contrast, the main block group-level file summarized in Section 3 contains food retailers for all block groups in

---

<sup>5</sup> The selection of block groups in which FoodAPS households reside was a way to approximate the block groups in SSUs. This approach was easier than simply selecting SSU block groups, because, as previously noted, the SSUs were sampled at a time when census 2000 block group boundaries were in use, while FoodAPS household locations are identified in terms of census 2010 block group boundaries.

each PSU county. There are two reasons for restricting the restaurant file to block groups in which FoodAPS households reside: (1) For four large PSU counties, not all restaurants in the county were included; data include only those restaurants located within 5 miles of the centroid of ZIP codes containing SSUs; and (2) information was not available for restaurants in counties adjacent to PSU counties, so data quality for the many non-SSU block groups on the boundaries of PSUs could not be ensured. Geographic analysis was restricted to restaurants within short distances (0.5 mile and 1 mile) of block group centroids to minimize boundary issues.

Restaurants comprised all establishments with a 4-digit InfoUSA SIC code of 5812 (Eating Places). Restaurants were classified as limited-service or full-service restaurants based on their six-digit primary InfoUSA SIC codes. If the primary SIC code was not an “Eating Place,” the six-digit secondary SIC code was used instead. In general, hamburger, pizza, sandwich, deli, and carryout establishments were classified as limited-service restaurants. Full-service restaurants included all other eating establishments. Restaurants excluded coffee shops; ice cream, snack, soft drink, and soda fountain establishments; caterers; and contract food services. Establishments with SIC code 581208 (Restaurants) were considered full-service restaurants by default, unless they belonged to a national or regional restaurant chain.<sup>6</sup> Chain restaurants were reassigned as necessary, using information on the type of service provided from company webpages.

The variables in this file follow the conventions described in the previous section. The prefix *N\_* indicates that these are counts of restaurants. The distances from the population-weighted block group centroid are indicated by a suffix to the variable name (Table 1). The middle part of the

---

<sup>6</sup> Restaurants with the same name appearing at least 20 times in the InfoUSA dataset were considered part of a restaurant chain.

variable name indicates the type of restaurant (Table 4). The meaning of the variable name comes from the prefix, the middle, and the suffix together. For example, for each block group, the variable *N\_LSRestaurant\_d1* records the non-negative integer number of limited-service restaurants located less than or equal to 0.5 miles from a block group centroid.

**Table 4.** Variable name descriptions for restaurants.

Name	Full Description
<b>LSRestaurant</b>	Limited-service restaurants. These establishments are primarily engaged in providing food services where patrons generally order, select, and pay for items before eating. Food and drink may be consumed on premises, taken out, or delivered to the customers' location. These include hamburger, pizza, sandwich, deli, and carryout establishments. Based on InfoUSA data on eating places (4-digit SIC code 5812).
<b>FSRestaurant</b>	Full-service restaurants. These establishments are primarily engaged in providing food services to patrons who order and are served while seated and pay after eating. Based on InfoUSA data on eating places (4-digit SIC code 5812).

Note: The files named BGRestaurant denote the most current version of the main block group file.

## 5. Census Tract- and County-Level Files

Where applicable, the variable names in the census tract-level file follow the conventions described for the block group files in Tables 1, 2, and 3. Additional conventions unique to the census tract-level are described in Tables 5 and 6. Furthermore, the individual variable names are described in the Variable List-GC.

Many variables describe counts of retailers or other food sources within selected distances from the population-weighted block group centroid, averaged across all block groups in each census tract. For instance, the variable *N\_Supermarket\_d1* in the census tract-level file is the population-weighted mean of the non-negative integer number of supermarkets/supercenters/large grocery stores located less than or equal to 0.5 miles from a block group centroid, across all block groups in the census tract. Other variables represent population or household counts at the block group level, aggregated to the census tract level. For instance, the variable *nocar* in the census tract-level file is the number of occupied housing units without access to a vehicle, aggregated to census tract level. Food access variables from the ERS Food Access Research Atlas (FARA) have been attached to the census tract-level file.

In the county-level file, the variables describing counts of retailers or other food sources are for the county as a whole and are not based on selected distances from the population-weighted block group centroids. The restaurant counts are based on information from the 2012 County Business Patterns rather than the InfoUSA database. Other variables representing population or household counts are for the county as a whole. Policy variables are typically State-level variables attached to each county.

Variables with labels beginning with “ERS” are from the ERS Food Access Research Atlas; “CCD” are from NCES Common Core of Data; “CLASS” are from NCI Classification of Laws Associated with School Students; “BTG” are from Bridging the Gap; and “SNAP” are from SNAP Policy Database.

**Table 5.** Variable name prefixes indicate the type of variable.

<b>Prefix</b>	<b>Full Description</b>
<b>LACTpop</b>	Tract sum, persons living in a low access block group
<b>LACThunv</b>	Tract sum, housing units without vehicle access with low access
<b>lilattracts</b>	Low income and low access tract (FARA)
<b>la or lattracts</b>	Low access tract (FARA)
<b>lapop</b>	People with low access (FARA)
<b>lalowi</b>	Low-income people with low access (FARA)
<b>lakids</b>	Children age 0-17 with low access (FARA)
<b>laseniors</b>	Seniors age 65+ with low access (FARA)
<b>lahunv</b>	Housing units without vehicle access and with low access (FARA)

**Table 6.** Variable name suffixes indicate the metric of the variable.

<b>Suffix</b>	<b>Full Description</b>
<b>1and10</b>	Attribute at 1 mile for urban areas and 10 miles for rural areas (FARA)
<b>1and20</b>	Attribute at 1 mile for urban areas and 20 miles for rural areas (FARA)
<b>halfand10</b>	Attribute at 0.5 mile for urban areas and 10 miles for rural areas (FARA)
<b>half</b>	Attribute or count at 0.5 mile (FARA)
<b>halfshare</b>	Share at 0.5 mile (FARA)
<b>1</b>	Attribute or count at 1 mile (FARA)
<b>1share</b>	Share at 1 mile (FARA)
<b>10</b>	Attribute or count at 10 miles (FARA)
<b>10share</b>	Share at 10 miles (FARA)
<b>20</b>	Attribute or count at 20 miles (FARA)
<b>20share</b>	Share at 20 miles (FARA)

Note: The file named Tract denotes the most current version of the tract-level file, and the file named County denotes the most current version of the county-level file.